Comparing Broadband and Mobile Traffic

2.1 Overview

In this report we analyze traffic over the broadband access services operated by IIJ every year and present the results[1][2][3][4][5][6]. We once again report on changes in traffic trends over the past year based on daily user traffic and usage by port. From this report, we will also report on traffic trends for mobile service traffic in the same way.

Figure 1 shows average monthly traffic trends for IIJ's overall broadband services and mobile services. The IN/OUT traffic indicates directions from an ISP's perspective. IN represents uploads from users, and OUT represents user downloads. Because we cannot disclose traffic numbers, the data for each is normalized to 1 for the maximum value.

For broadband, over the past year IN traffic has increased by 15%, while OUT traffic has increased by 38%. Over the last eight years download volumes have grown significantly more than upload volumes, with IN traffic increasing by factor of 1.6, and OUT traffic increasing by a factor of 3.8.

For mobile we only present figures for the past year, but over this period traffic grew rapidly, with IN traffic increasing by a factor of 3.5 and OUT traffic increasing by a factor of 4.2. However, the total volume of mobile traffic is still an order of magnitude lower than broadband.

2.2 About the Data

For broadband traffic, as with our previous reports, the survey data utilized here was collected using Sampled NetFlow from the routers accommodating fiber-optic and DSL broadband customers of our personal and enterprise broadband access services. For mobile traffic, access gateway billing information was used to determine usage amounts for personal and enterprise mobile services, while Sampled NetFlow data from the routers accommodating these services was employed to determine the ports used.

Because traffic trends differ between weekdays and weekends, we analyze a full week of traffic. In this case, we have used data for the week spanning June 1 to June 7, 2015. For broadband traffic we compare results with those for the week spanning May 26 to June 1, 2014, which we analyzed in the previous report.

Figure 1: Trends in Monthly Traffic for Broadband and Mobile
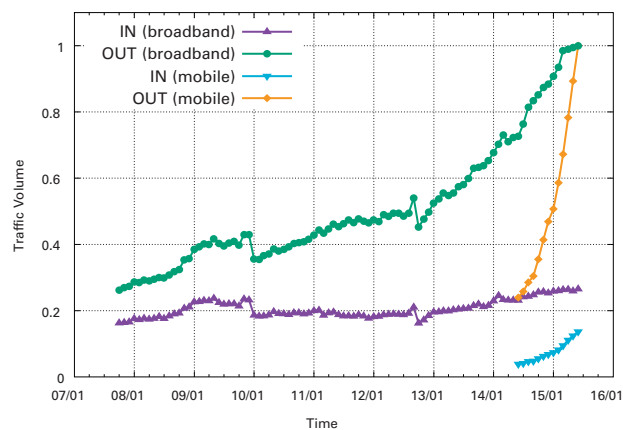
Results are aggregated by subscription for broadband traffic, and by phone number for mobile traffic, as some subscriptions cover a number of phone numbers. The usage volume for each broadband user was obtained by matching the IP address assigned to users with the IP addresses observed. We collected statistical information by sampling packets using NetFlow. The sampling rate was set between 1/8192 and 1/16384, taking into account router performance and load. We estimated overall usage volumes by multiplying observed volumes by the reciprocal of the sampling rate. IIJ provides both fiber-optic and DSL access for its broadband services. However, fiber-

*1   Kenjiro Cho. Broadband Traffic Report: Traffic Volumes Rise Steadily Over the Past Year, and HTTPS Use Expands. Internet Infrastructure Review. Vol.24. pp28-33. August 2014.
*2   Kenjiro Cho. Broadband Traffic Report: The Impact of Criminalization of Illegal Downloads was Limited. Internet Infrastructure Review. Vol.20. pp32-37. August 2013.
*3   Kenjiro Cho. Broadband Traffic Report: Traffic Trends over the Past Year. Internet Infrastructure Review. Vol.16. pp33-37. August 2012.
*4   Kenjiro Cho. Broadband Traffic Report: Examining the Impact of the Earthquake on Traffic on a Macro Level. Internet Infrastructure Review. Vol.12. pp25-30. August 2011.
*5   Kenjiro Cho. Broadband Traffic Report: Traffic Shifting away from P2P File Sharing to Web Services. Internet Infrastructure Review. Vol.8. pp25-30. August 2010.
*6   Kenjiro Cho. Broadband Traffic: Increasing Traffic for General Users. Internet Infrastructure Review. Vol.4. pp 18-23. August 2009.

optic access now makes up the vast majority of use, with 95% of users observed in 2015 using fiber-optic connections, accounting for 98% of overall broadband traffic volumes.

2.3 Daily Usage Levels for Users

First, we will examine the daily usage volumes for broadband and mobile users from several perspectives. Daily usage indicates the average daily usage calculated from a week's worth of data for each user.

Figure 2 and Figure 3 show the average daily usage distribution (probability density function) per broadband and mobile user. They compare data for 2014 and 2015 divided into IN (upload) and OUT (download), with user traffic volume on the X axis, and user frequency on the Y axis.The X axis shows volumes between 10 KB (10^4) and 100 GB (10^{11}) using a logarithmic scale. Some users are outside the scope of the graph, but most fall within the 100 GB (10^{11}) range.

The IN and OUT distribution for broadband shows almost log-normal distribution, which looks like a normal distribution in a semi-log graph. A linear graph would show a long-tailed distribution, with the peak close to the left end and a slow decay towards the right. The OUT distribution is further to the right than the IN distribution, indicating that the download volume is more than an order of magnitude larger than the upload volume. Comparing 2014 and 2015, the peak distribution for both IN and OUT traffic has moved slightly to the right, demonstrating that overall user traffic volumes are increasing. Looking at OUT distribution, the peak has been steadily moving to the right over the past few years. However, the usage levels of heavy users on the right end have not increased much, and the distribution is beginning to lose its symmetry. Meanwhile, the tail of the IN distribution to the right has grown longer. Previously, both IN and OUT showed a clearer peak here, indicating heavy users with symmetrical IN/OUT volumes. For convenience, we labeled users with asymmetrical IN/OUT traffic distribution that make up the majority "client-type users," and the distribution of heavy users with symmetrical IN/OUT traffic that make up the minority on the right side "peer-type users." In this report we will continue to use these conventions. Over the past few years, the peak for peer-type users has shrunk to the point where it can hardly be distinguished. This indicates that the ratio of heavy users is decreasing.

The data for mobile traffic in Figure 3 indicates that usage volumes are significantly lower than broadband. Additionally, because there are limits on data usage, the ratio of heavy users to the right of the distribution is lower, creating left-right asymmetry. There are also no extremely heavy users. Due to those who use mobile only when going out and limits on data usage, there is greater variance in daily usage volumes for each user compared to broadband. For this reason, when you look at the daily average for a week worth of data, there is less variance between users than when examining individual days. Plotting distribution for individual days in the same way results in lower peaks, and raises the tails on either side, but the basic form and most frequent values of the distribution remain largely unchanged.
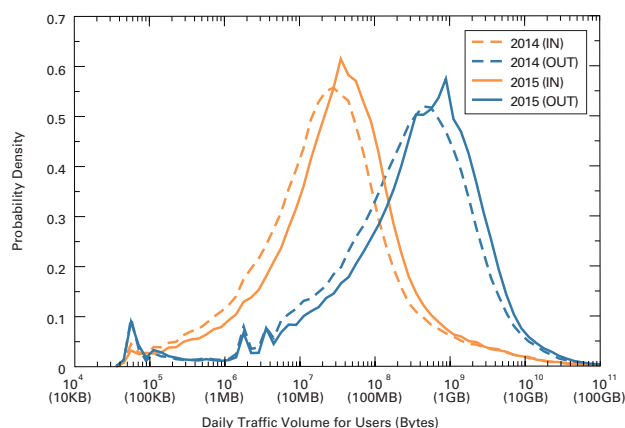

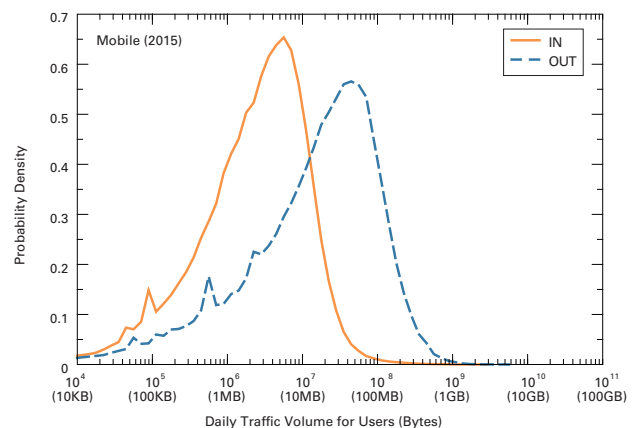Figure 2: Daily Broadband User Traffic Volume Distribution Comparison of 2014 and 2015

Figure 3: Daily Mobile User Traffic Volume Distribution in 2015

Table 1 shows trends in the mean daily traffic value for broadband users as well as the modal value (the most frequent value that represents peak distribution). This year the peak position has shifted slightly from the center of the distribution so that the modal value was corrected towards the center of the distribution. Comparing the modal values in 2014 and in 2015, IN rose from 28 MB to 40 MB, and OUT rose from 447 MB to 708 MB. This demonstrates that, particularly in the case of downloads, the traffic volume for each user has increased. Meanwhile, because mean values are pulled up by the heavy users to the right of the graph, they are significantly higher than the modal values, with the mean IN value 467 MB and the mean OUT value 1,621 MB in 2015. In 2014, these were 437 MB and 1,287 MB, respectively. For mobile, as shown in Table 2, there is little difference between the mean and modal values due to the lack of extremely heavy users. The modal values were 5.5 MB for IN and 40 MB for OUT, with the mean values amounting to 6.0 MB for IN and 46.6 MB for OUT.

Figure 4 and Figure 5 plot the IN/OUT usage volumes for 5,000 randomly sampled users. The X axis shows OUT (download volume) and the Y axis shows IN (upload volume), with both using a logarithmic scale. Users with identical IN/OUT values are plotted on the diagonal line.

The cluster below the diagonal line and spread out parallel to it represents general client-type users with download volumes an order of magnitude higher than upload volumes. For broadband traffic, there was previously a clearly-recognizable cluster of peer-type heavy users spread out thinly on the upper right of the diagonal line, but this is now no longer discernible. Though we have separated client-type and peer-type users for convenience, in actual fact client-type general users also use peer-type applications such as Skype, and peer-type heavy users also use download-based applications on the Web, blurring the boundary between them. In other words, many users use both types of applications in varying ratios. There are also differences in the usage levels and IN/OUT ratio for each user, pointing to the existence of diverse forms of usage. Here, almost no difference can be discerned compared to 2014.

Year	IN (MB/day)		OUT (MB/day)	
	Mean	Mode	Mean	Mode
2005	430	3.5	447	32
2007	433	4	712	66
2008	483	5	797	94
2009	556	6	971	114
2010	469	7	910	145
2011	432	8.5	1,001	223
2012	410	14	1,026	282
2013	397	18	1,038	355
2014	437	28	1,287	447
2015	467	40	1,621	708

Table 1: Trends in Mean and Mode of Daily Traffic Volume for Each Broadband User

Year	IN (MB/day)		OUT (MB/day)	
	Mean	Mode	Mean	Mode
2015	6.0	5.5	46.6	40

Table 2: Trends in Mean and Mode of Daily Traffic Volume for Each Mobile User
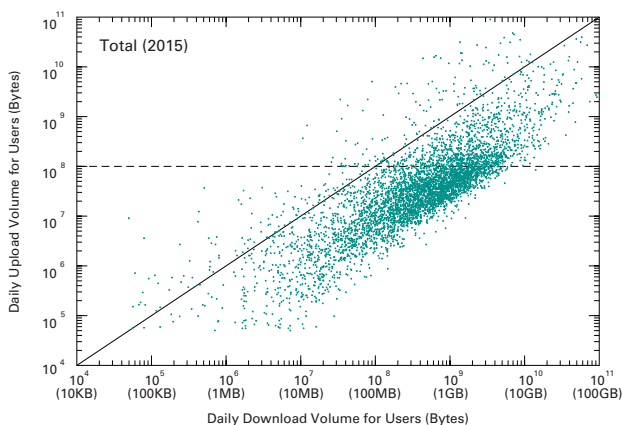
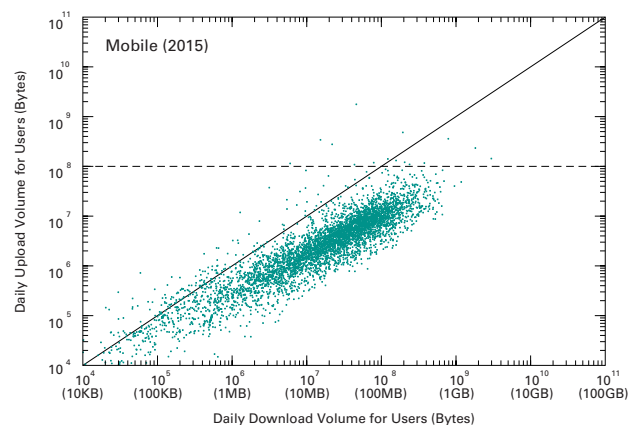

Figure 4: IN/OUT Usage for Each Broadband User

Figure 5: IN/OUT Usage for Each Mobile User

The trend for OUT traffic to be an order of magnitude larger also applies to mobile, but usage volumes are lower than broadband, and there is less variance between IN/OUT.

Figure 6 and Figure 7 show the complementary cumulative distribution of the daily traffic volume for users. This indicates the percentage of users with daily usage levels greater than the X axis value on the Y axis in a log-log scale, which is an effective way of examining the distribution of heavy users. The right side of the graph falls linearly, showing a long-tailed distribution close to power-law distribution. It can be said that heavy users are distributed statistically, and are by no means a special class of user. Heavy users also exhibit power-law distribution for mobile, but the ratio of heavy users is lower, and they also have OUT usage levels several times higher than their IN amounts.

There is a great deal of deviation in traffic usage levels between users, and as a result traffic volume for a small portion of users accounts for the majority of overall traffic. For example, the top 10% of broadband users make up 65% of the total OUT traffic, and 90% of the total IN traffic. Furthermore, the top 1% of users make up 29% of the total OUT traffic, and 63% of the total IN traffic. Along with the decrease in the ratio of heavy users over the past few years, the distribution bias is also dropping slightly.

For mobile, the top 10% of users account for 49% of OUT traffic and 50% of IN traffic, while the top 1% make up 13% of OUT traffic and 19% of IN traffic. This also demonstrates the low ratio of heavy users among mobile users.

2.4 Usage by Port

Next, we will look at a breakdown of traffic and examine usage levels by port. Recently, it has been difficult to identify applications by port number. Many P2P applications use dynamic ports on both ends, and a large number of client/server applications utilize port 80 assigned to HTTP to avoid firewalls. To broadly categorize, when both parties use a dynamic port higher than port 1024, there is a high possibility of it being a P2P application, and when one party uses a well-known port lower than port 1024, it is likely to be a client/server application. In light of this, here we will look at usage levels for TCP and UDP connections by taking the lower port number of the source and destination ports.
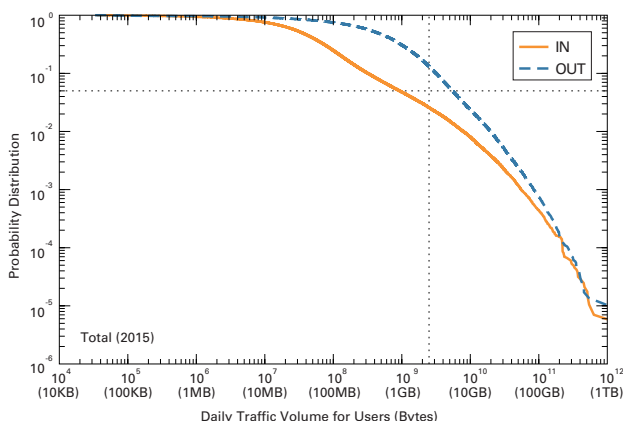
Figure 6: Complementary Cumulative Distribution of the Daily Traffic Volume for Broadband Users
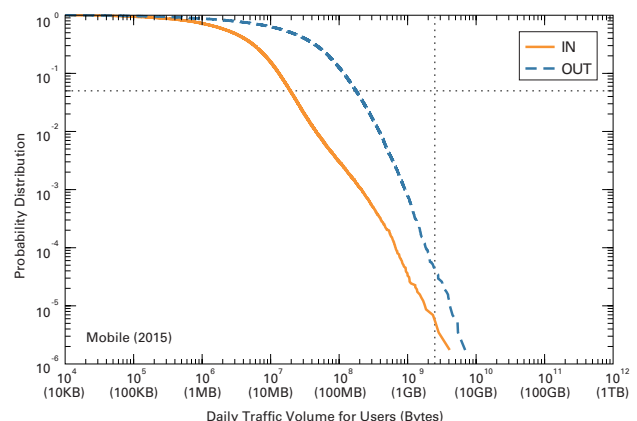
Figure 7: Complementary Cumulative Distribution of the Daily Traffic Volume for Mobile Users

As overall traffic is dominated by peer-type heavy user traffic, to examine trends for client-type general users, we have taken the rough approach of extracting data for users with a daily upload volume of less than 100 MB, and treating them as client-type users. This corresponds to users below the horizontal line at the IN=100 MB point in Figure 4.

Table 3 compares port usage ratios for broadband users in 2014 and 2015, showing total users and client-type users.

81% of overall traffic in 2015 is TCP based. The ratio of port 80 (HTTP) traffic was down from 45% in 2014 to 38% this year. The ratio of port 443 (HTTPS) traffic climbed from 9% to 23%. TCP dynamic port traffic, which has been on the decline, dropped from 24% in 2014 to 18% in 2015. The ratio of individual dynamic port numbers is tiny, with port 1935 used by Flash Player the highest at 2% of the total, and the next highest under 0.5%. Aside from TCP, there is port 443 (HTTPS) UDP traffic, which is thought to be Google's QUIC protocol. The rest is mostly VPN-related.

Looking exclusively at client-type users, port 80 traffic that accounted for 75% of the total in 2014 dropped 22 percentage points to 53% in 2015. Meanwhile, the next highest traffic ratio for port 443 (HTTPS) traffic climbed 21 percentage points, going from 14% in 2014 to 35% in 2015. We believe this was due to the shift of some HTTP traffic to HTTPS. The ratio of dynamic ports also decreased from 7% to 5%.

Table 4 shows port usage ratios for mobile users, which as a whole were close to the values for client-type users on broadband.

The growth in the use of HTTPS is due to more and more services making regular use of encrypted HTTPS communications since the existence of a U.S. National Security Agency (NSA) program for intercepting communications stirred up controversy in June 2013. Looking at HTTPS traffic volumes broken down by provider for 2015, about 80% was related to Google, demonstrating their efforts to proactively adopt HTTPS. We also believe YouTube traffic volumes are boosting HTTPS usage.

protocol port	2014		2015	
	total (%)	client type	total (%)	client type
TCP	**80.2**	**97.4**	**80.8**	**94.7**
(< 1024)	*56.3*	*90.1*	*63.3*	*89.9*
80 (http)	44.9	74.8	37.9	53.2
443 (https)	9.3	13.8	23.3	35.1
81	0.0	0.0	0.5	0.7
182	0.0	0.0	0.4	0.3
554 (rtmp)	0.4	0.3	0.2	0.1
22 (ssh)	0.3	0.0	0.2	0.0
(>= 1024)	*23.8*	*7.3*	*17.5*	*4.8*
1935 (rtmp)	2.5	4.0	1.8	2.4
8080	0.4	0.2	0.3	0.1
7144 (peercast)	0.3	0.0	0.2	0.0
UDP	**12.5**	**1.8**	**11.4**	**2.6**
443 (https)	0.0	0.0	0.9	0.9
1701 (12tp)	0.0	0.0	0.2	0.3
ESP	**6.9**	**0.7**	**7.4**	**2.6**
IP-ENCAP	**0.2**	**0.0**	**0.2**	**0.0**
GRE	**0.2**	**0.0**	**0.2**	**0.0**
ICMP	**0.0**	**0.0**	**0.0**	**0.0**

Table 3: Usage by Port for Broadband Users

protocol port	2015
	total (%)
TCP	**93.8**
80 (http)	52.5
443 (https)	37.4
81	0.5
993 (imaps)	0.5
1935 (rtmp)	0.5
UDP	**5.2**
1701 (12tp)	1.8
443 (https)	1.0
4500 (nat traveral)	0.3
80 (http)	0.2
53 (dns)	0.1
ESP	**0.7**
GRE	**0.3**
ICMP	**0.0**

Table 4: Usage by Port for Mobile Users

Figure 8 compares trends in TCP port usage over a week for overall broadband traffic in 2014 and 2015. Trends in TCP port usage are shown for four categories: port 80, port 443, other well-known ports, and dynamic ports. Traffic is normalized to 1 for the total peak traffic volume. Previously we included port 443 under other well-known ports, but due to the increase in its usage ratio, from this report we are plotting port 443 separately. Compared with 2014, we can see that the overall ratio of port 443 usage has increased further, and the use of dynamic ports is decreasing. The overall peak is between 21:00 and 1:00 the next day, and traffic also increases in the daytime on Saturday and Sunday, reflecting times when the Internet is used at home.

Figure 9 shows trends for port 80 and port 443 with regard to mobile, where traffic remains high from morning to night. On weekdays there are three peaks representing the morning commute hours, lunch break, and evening to night periods, demonstrating that usage times are different from broadband.

2.5 Conclusion

One trend we have seen in broadband traffic over the past year is significant growth in HTTPS usage, which we reported on in our last report. Other than this there were no major changes in overall trends, although this year download volumes increased 38%, and upload volumes climbed 15%. A year before these increases were 27% and 13%, so over the past year the rate of increase has also gone up.

Although still falling far short of broadband volumes, the mobile traffic we began reporting on in this volume has grown significantly in the past year. Due to bandwidth and data cap limitations, there are less heavy users for mobile compared to broadband. Usage ratios by port are close to light broadband users, suggesting the use of the same kind of services as broadband on mobile due to the increasing popularity of cloud-based services accessible from multiple devices. Meanwhile, some points were different to broadband, such as the standout usage times being the commuting hours and lunch breaks on weekdays.
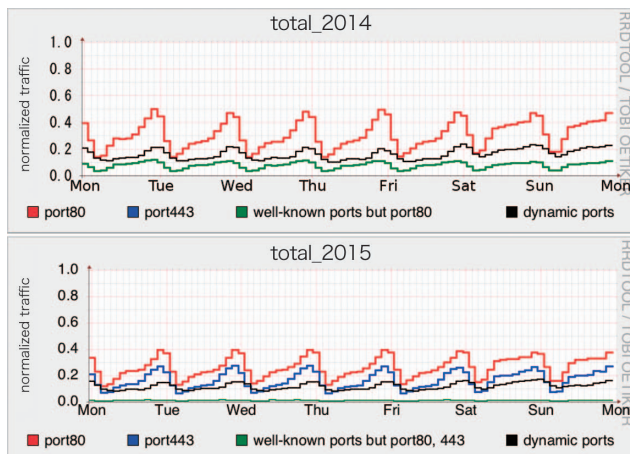
Figure 8: Weekly TCP Port Usage Trends for Broadband Users in 2014 (top) and 2015 (bottom)
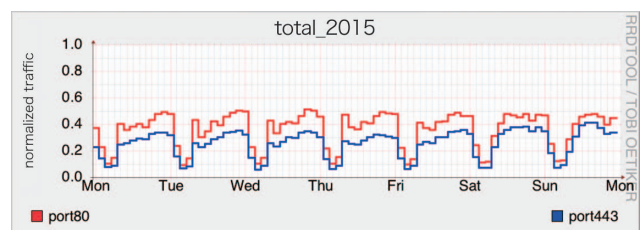
Figure 9: Weekly TCP Port Usage Trends for Mobile Users in 2015

Author:
Kenjiro Cho
Research Director, Research Laboratory, IIJ Innovation Institute Inc.